

Sveučilište Josipa Jurja Strossmayera u Osijeku

Filozofski fakultet

Informatologija

Ivona Grgić

Od UNIMARCA do povezanih podataka i semantičkog weba

Završni rad

Mentor prof.dr.sc. Kornelija Petr Balog

Osijek, 2017.

Sveučilište Josipa Jurja Strossmayera u Osijeku

Filozofski fakultet

Odsjek za informacijske znanosti

Informatologija

Ivona Grgić

Od UNIMARCA do povezanih podataka i semantičkog weba

Završni rad

Područje: društvene znanosti; Polje: informacijske i komunikacijske znanosti;
Grana: knjižničarstvo

Mentor prof.dr.sc. Kornelija Petr Balog

Osijek, 2017.

Sadržaj

Sažetak	4
Ključne riječi	4
1. Uvod.....	5
2. Universal Machine Readable Cataloging (UNIMARC)	6
2.1. Nastanak i razvoj formata UNIMARC	6
2.2. Namjena i opseg UNIMARC-a	8
3. Općenito o semantičkom webu i povezanim podacima.....	9
3.1. Temeljna načela semantičkog weba i povezanih podataka	12
4. Bibliografski podaci kao povezani podaci	12
4.1. Knjižnični zapis UNIMARC u okruženju semantičkog weba.....	13
4.2. Problemi primjene koncepta povezanih podataka na knjižničnim zapisima	19
Zaključak	21
Literatura	22

Sažetak

Ovaj rad prikazuje knjižnični zapis koji je strojno čitljiv na formatu UNIMARC te koji postaje strojno „razumljiv“ u okruženju povezanih podataka, odnosno semantičkog weba. Na početku rada govori se o povijesti i razvoju formata UNIMARC te koja je njegova glavna svrha i namjena u svijetu bibliografskih podataka. Nadalje, govori se općenito o konceptu povezanih podataka i semantičkog weba te uz definiranje koncepta navode se i objašnjavaju načela na kojima počivaju. Definirane su i glave sastavnice povezanih podataka – URI (Uniform Resource Identifiers) i RDF (Resource Description Framework). U dijelu rada gdje su bibliografski podaci prikazani u okruženju semantičkog weba najprije se navode izazovi s kojim se susreću svi koji sudjeluju u prikazivanju strojno čitljivih bibliografskih podataka kao povezanih podataka. Nakon toga, prikazuje se, na primjeru iz knjižničnog kataloga CROLIST, semantičko modeliranje bibliografskog zapisa u UNIMARC formatu kroz tri koraka: odabir, identificiranje i objavljivanje zapisa. Na kraju, navode se glavni problemi objave knjižničnih bibliografskih podataka kao povezanih podataka. Uz tehničke i organizacijske probleme kao najveća zapreka navodi se nepostojanje jedinstvenog rječnika i imenskih prostora koji su vrlo važni za identificiranje pojedinih bibliografskih izvora.

Ključne riječi: UNIMARC, povezani podaci, semantički web, RDF, URI

1. Uvod

Cilj ovoga rada je predstaviti kako je bibliografski zapis nastao strojno čitljiv na međunarodnom formatu te kako postaje strojno „razumljiv“ u okruženju povezanih podataka. Rad je podijeljen na tri poglavlja, tako da se u prvom poglavlju definira format UNIMARC te se govori o nastanku, razvoju i namjeni formata. Također, navodi se kako se format razvijao u pojedinim zemljama svijeta i kako se došlo do međunarodnog formata za razmjenu bibliografskih zapisa. Nadalje, drugi dio se odnosi na općeniti pogled semantičkog weba i povezanih podataka te na kojim se načelima oni temelje. U zadnjem dijelu prikazuje se važnost prikazivanja bibliografskih podataka kao povezanih podataka te s kojim se izazovima susreću knjižnice kako bi ih prikazale kao takve. Zatim se kroz primjer iz knjižničnog kataloga predstavlja semantičko modeliranje u tri koraka odnosno prikazuje se knjižnični zapis u formatu UNIMARC u okruženju semantičkog weba. Naravno, kako i svaki prijelaz na novo okruženje tako i ovaj ima poteškoće s kojima se susreće te su na kraju predstavljeni glavni problemi s kojima se susreću svi koji sudjeluju u semantičkom modeliranju bibliografskih podataka.

Katalogiziranje se odnosi na opisivanje (koristeći standarde i kontrolirane rječnike) izvora i omogućavanje njihove dostupnosti. To su metapodaci, s kojima smo dugo zaglavili u svijetu MARC-a. Dok se internet, s vlastitim standardima i rječnicima, razvio oko nas. Knjižnica ima puno toga za ponuditi internetu i bila bi pogreška da internet ignorira knjižnicu, kao i knjižnica internet.¹ Zbog toga se tradicionalna struktura MARC-a i arhitektura relacijskih baza podataka zamjenjuje novim okruženjem povezanih podataka na semantičkom webu. Znanje i razumijevanje katalogizatora i metapodatkovnih specijalista o MARC-u i pravilima AACR-a, koje su koristili u prošlosti, vrlo je važno za prijelaz iz zatvorenog okruženja u ekspanzivni svijet povezanih podataka odnosno prijelaz u budućnost.² Značajka povezanih podataka je njegov semantički web koji omogućuje dijeljenje povezanih podataka na globalnoj razini između različitih izvora te bi on trebao biti od velike važnosti za knjižnice koje razmjenjuju zapise formatom MARC još od 1960-ih godina.

¹ Usp. Frederick, E. Donna. Metadata specialists in transition: from MARC cataloging to linked data and BIBFRAME (data deluge column). // Library Hi Tech News 33, 4(2016), str. 3. URL: <http://www.emeraldinsight.com/doi/abs/10.1108/LHTN-03-2016-0015> (2017-07-08)

² Usp. Isto. Str. 3-4.

2. Universal Machine Readable Cataloguing (UNIMARC)

MARC (Machine Readable Cataloguing) odnosno strojno čitljivo katalogiziranje - opisuje se kao skupina formata koji primjenjuju određeni skup konvencija za identifikaciju i uređivanje bibliografskih podataka radi obrade računalom.³ Međunarodna federacija knjižničkih društava i ustanova - IFLA (International Federation of Library Associations and Institutions) pokrenula je realizaciju programa Univerzalne bibliografske kontrole (UBC – Universal Bibliographic Control) koji se zasnivao na načelima da bi svaka zemlja trebala preuzeti odgovornost za bibliografsku kontrolu vlastitih publikacija i da zapise koje izradi jedna nacionalna ustanova treba staviti na raspolaganje drugim zemljama za katalogizaciju istih publikacija.⁴ U tu je svrhu IFLA razvila format UNIMARC: međunarodni standard za prikaz i razmjenu bibliografskih podataka.⁵ Definicija tog formata ujedno predstavlja i njegovu osnovnu svrhu, dakle, da omogući međunarodnu razmjenu bibliografskih podataka u strojno čitljivom obliku između nacionalnih bibliografskih ustanova.⁶

2.1. Nastanak i razvoj formata UNIMARC

Potreba ispitivanja elemenata i strukture bibliografskih podataka te postojećih pravila za njihovo organiziranje pri izradi kataloga na listićima pojavljuje se kao posljedica uvođenja računala u knjižnice. Jedinstveno opisivanje, organiziranje u skupove elemenata podataka te određivanje međusobnih odnosa bilo je potrebno za identificirane bibliografske podatke i njihovu strukturu kako bi im se dodijelili standardizirani kodovi u okviru formata za strojno čitljivo katalogiziranje.⁷ Sredinom šezdesetih godina Kongresna je knjižnica u SAD-u pokrenula istraživanje usmjereno na rješavanje osnovnih zadataka: razviti format za strojno čitljivo bilježenje bibliografskih zapisa, istražiti sadašnje i buduće upotrebe bibliografskih zapisa u automatiziranoj sredini i istražiti načine na koje strojno čitljiv format može omogućiti prenošenje i razmjenu kataložnih zapisa između knjižnica.⁸ Istraživanje je uzrokovalo pojavu prvog MARC formata – MARC I. Paralelno se razvijao MARC format u SAD-u i Velikoj Britaniji te su suradnjom Kongresne knjižnice i Savjeta Britanske nacionalne bibliografije nastala dva srodna formata: USMARC (United States MARC) i UKMARC (United Kingdom MARC). Nadalje, razvijaju se CANMARC (Kanada), AUSMARC (Australija), ANNAMARC (Italija), IBERMARC (Španjolska), DANMARC (Danska) itd. 1973.

³ Usp. Willer, Mirna. UNIMARC u teoriji i praksi. Rijeka: Naklada „Benja“, 1996. Str. 13.

⁴ Usp. Isto. Str. 13-14.

⁵ Usp. UNIMARC. // Hrvatska enciklopedija. Leksikografski zavod Miroslav Krleža. URL: <http://www.enciklopedija.hr/natuknica.aspx?ID=63215> (2017-04-26)

⁶ Usp. UNIMARC: bibliografski format. Zagreb: Hrvatsko knjižničko društvo; Zadar: Sveučilište, 2009. Str. 6.

⁷ Usp. Willer, Mirna. Nav. dj. Str. 14.

⁸ Usp. Isto.

Njemačka je objavila svoj nacionalni format – MAB (Maschinelles Austauschformat für Bibliotheken) te je njegova tehnika povezivanja zapisa uvelike utjecala na kasniji razvoj formata UNIMARC. No obilježje prvog međunarodnog MARC formata može se pripisati INTERMARC-u (International MARC Format) koji je objavljen 1975. godine, ali on je zadržan kao interni format Nacionalne knjižnice Francuske, dok se UNIMARC koristi kao format za razmjenu.⁹ MARC je omogućio prvobitnu upotrebu računalne tehnologije u knjižnicama te se sastoji od različitih polja, također nazvanih identifikatori ili oznake, koja su povezana s informacijama o bibliografskom entitetu. Postoji puno mogućih polja, uključujući osnovna kao što su navedena u nastavku¹⁰:

010 LC kontrolni broj

020 ISBD

100 Odrednica – osobno ime

245 podaci o naslovu, podnaslovu i odgovornosti

250 izdanje

300 materijalni opis

440 podatak o nakladničkoj cjelini

500 napomena

650 tematska predmetna odrednica

700 sporedna kataložna jedinica.

Početkom 70-ih godina započeo je rad na UNIMARC-u te osnova za njegovu izradu poslužili su standardi: Pariška načela, međunarodni standardni bibliografski opis i međunarodni format za razmjenu bibliografskih podataka na magnetskom mediju. Tim standardima rješavali su se osnovni problemi poput odabira i oblika odrednica, kataložnog opisa i strukture formata za razmjenu bibliografskih podataka, no kao neriješeni problem ostaje pomanjkanje standardiziranih označitelja sadržaja strojno čitljivih bibliografskih zapisa. 1972. IFLA je osnovala Radnu grupu za označitelje sadržaja koja je bila zadužena za određivanje označitelja sadržaja za sve vrste građe, međutim, Radna grupa se odlučila za prvi nacrt formata uključiti samo onu vrstu građe koja je opisana ISBD-om (međunarodni standardni bibliografski opis) te je također ograničila područje

⁹ Usp. Isto. Str. 14-16.

¹⁰ Usp. Rubin, E. Richard. Foundations of Library and Information Science. New York; London: Neal-Schuman Publishers, 2004. Str. 245-246.

svog rada na potrebe knjižnica i nacionalnih bibliografija.¹¹ Prvo izdanje UNIMARC-a IFLA objavljuje 1977. godine te objašnjava kako je sam format kreiran i za koju je svrhu namijenjen.

„Format je napravljen tako da svaka nacionalna bibliografska ustanova odgovara za prevođenje bibliografskih zapisa sa svoga nacionalnog formata u UNIMARC radi njihove razmjene s drugim nacionalnim ustanovama, te za prihvaćanje strojno čitljivih zapisa u formatu UNIMARC od drugih nacionalnih ustanova i njihovo prevođenje u nacionalni format.“¹²

Kao što se može vidjeti iz navedenog citata pojedina nacionalna bibliografska ustanova odgovorna je za prevođenje svog formata u UNIMARC kako bi lakše razmjenjivala zapise s drugim nacionalnim ustanovama i obrnuto. Nadalje, godine 1980. IFLA je objavila drugo, prerađeno izdanje koje uključuje polja podataka za kartografsku građu te za omeđene i serijske publikacije. Nadalje, trećim izdanjem smatra se UNIMARC Manual, objavljen 1987., koji sadrži promjene i dopune u poljima u skladu sa zadnjim izdanjima ISBD-a i uključuje nova polja: polja kodiranih podataka i napomena za zvučne snimke i muzikalije, polja za bilježenje koda nacionalnog ureda za ISDS, dodatna polja za kartografsku građu i privremena polja za grafiku i kompjutorske datoteke.¹³ Upotreba UNIMARC-a i potreba za približavanjem cilju UBC-a na području podataka preglednih kataložnih jedinica uzrokovali su objavljivanjem 1991. *UNIMARC-a za pregledne kataložne jedinice i uputnice*.¹⁴ 1994. izlazi četvrto izdanje pod nazivom UNIMARC Manual: bibliographic format, koje donosi nova polja za identifikator sastavnice, polja kodiranih podataka za trodimenzionalne umjetnine i realije, svojstva kompjutorskih datoteka, napomene i polja za povezivanje kataložnih jedinica za mikrooblike, polje za povezivanje jedinice koja se recenzira, stvarni naslov djela i privremeno polje za predmetnu grupu. 1996. godine izlazi prvo osuvremenjeno izdanje i donosi nova polja za opis starih omeđenih publikacija.¹⁵

2.2. Namjena i opseg UNIMARC-a

Namjena formata UNIMARC je da posluži kao format prijenosnik u svrhu razmjene. On ne propisuje oblik, strukturu ili sadržaj zapisa podataka unutar pojedinih sustava, nego donosi preporuke u sadržaju i obliku podataka kada se razmjenjuju. Onim ustanovama koje planiraju upotrebu UNIMARC-a za razmjenu podataka, korisno je ujednačiti označitelje sadržaja te definicije polja i potpolja unutar svoga sustava s onim u UNIMARC-u kako bi pojednostavili

¹¹ Usp. Willer, Mirna. Nav. dj. Str. 24-25.

¹² Usp. Isto. Str. 16.

¹³ Usp. Isto. Str. 26-27.

¹⁴ Usp. Isto. Str. 28.

¹⁵ Isto.

prevođenje podataka u trenutku kada se zapisi prevode u strukturu vrpce za razmjenu podataka.¹⁶ Ukratko rečeno, namjena UNIMARC formata je poticanje razmjene bibliografskih podataka u strojno čitljivome mediju između nacionalnih bibliografskih ustanova koje su odgovorne za prevođenje bibliografskih zapisa iz svoga nacionalnog formata u UNIMARC radi njihove razmjene s drugim ustanovama i njihova prevođenja u nacionalni format.¹⁷ Zatim, navedeno je što se formatu treba odrediti i za koje vrste publikacija uključuje označitelje sadržaja.

„Opseg je formata UNIMARC odrediti označitelje sadržaja (oznake polja, indikatore i oznake potpolja) koji se pridaju bibliografskih zapisima u strojnom čitljivom obliku, te odrediti logički i materijalni format zapisa. Format uključuje označitelje sadržaja za omeđene publikacije, serijske publikacije i drugu neomeđenu građu, kartografsku građu, muzikalije, zvučne snimke, grafiku, projiciranu filmsku građu i videograđu, staru i rijetku građu i arhivsku građu te elektroničku građu.“¹⁸

Što se tiče opsega formata UNIMARC za pregledne kataložne jedinice i uputnice, kao što je definirano u navedenom citatu, on određuje oznake polja, indikatore i oznake potpolja koji se pridaju zapisima preglednih kataložnih jedinica, općih uputnica i objasnidbenih uputnica u strojno čitljivom obliku. Oznake polja se odnose na osobna i obiteljska imena, nazive korporativnih tijela i sastanaka, nazive političkoteritorijalnih jedinica i zemljopisne nazive, jedinstvene stvarne naslove publikacija (anonimne) i jedinstvene stvarne naslove za djela individualnih autora, te za teme publikacije.¹⁹

3. Općenito o semantičkom webu i povezanim podacima

Semantički web definira se kao nadogradnja postojeće mreže u kojoj se informacijama dodjeljuje dobro definirano značenje koje omogućuje bolju suradnju računala i ljudi uz korištenje formalnih modela i primjenu univerzalnosti.²⁰ Tim Berners-Lee promatra semantički web kao mrežu “strojno čitljivih informacija čije je značenje dobro definirano standardima”. Cilj semantičkog weba je upotrijebiti odgovarajuću tehnologiju kako bi računala mogla pristupati, dijeliti i obraditi podatke iz različitih aplikacija te kako bi mogla razviti jezike koji će se moći upotrijebiti za izražavanje računalno obradivih informacija. Dakle, o njemu se može razmišljati kao o globalnoj bazi povezanih podataka.²¹

¹⁶ Usp. UNIMARC: bibliografski format. Nav. dj. Str. 7.

¹⁷ Usp. Willer, Mirna. Nav. dj. Str. 103.

¹⁸ UNIMARC: bibliografski format. Nav. dj. Str. 7.

¹⁹ Usp. Willer, Mirna. Nav. dj. Str. 137.

²⁰ Farago, Filip.; Bosančić, Boris.; Badurina, Boris. Povezani podaci i knjižnice. // Vjesnik bibliotekara Hrvatske 56, 4(2013), str. 29. URL: <http://hrcak.srce.hr/142376> (2017-03-25)

²¹ Usp. Sarić, Ivana. Model povezivanja i interoperabilnosti knjižničnih zapisa na semantičkom webu. // Vjesnik bibliotekara Hrvatske 56, 1/2(2013), str. 107-122. URL: <http://hrcak.srce.hr/file/170267> (2017-04-26)

W3 Konzorcij definirao je šest temeljnih načela na kojima se semantički web temelji:²²

1. Sve može biti identificirano URI-jem²³ - sve objekte na mreži moguće je identificirati pomoću sustava identifikatora URI te svatko tko ima nadzor nad dijelom mreže može dodjeljivati URI-je i određivati na što se oni odnose u fizičkom svijetu.

2. Izvori i poveznice mogu imati tipove - izvori u semantičkom webu postaju strojno razumljivi koncepti. Digitalni objekti koji predstavljaju predmete iz stvarnog svijeta postaju “klase”, a odnosi među njima nadalje modeliraju se pomoću RDF-a.

3. Djelomične informacije se dopuštaju - današnja mreža dopušta poveznice koje ne vode nikamo čime je žrtvovala integritet za skalabilnost, tj. stalan ili neočekivan rast. Semantički web također dopušta djelomične informacije, te je predviđeno da njegova pomagala mogu funkcionirati unatoč tome.

4. Nema potrebe za apsolutnom istinom - unutar semantičkog weba, svaka aplikacija zasebno zaključuje što je istina na osnovi dostupnih podataka. Pragmatično gledano, ona ne odlučuje o tome što je istina, nego kome će više vjerovati. Primjerice, podaci uglednih knjižnica mogu u nekom sustavu biti bolje rangirani od podataka neke manje poznate ‘treće strane’.

5. Razvoj se podržava - pretpostavljen je stav da je okupljanje sličnih koncepata iz različitih izvora korisno. Cilj je opisati izvore tako da dodavanje novijih podataka ne traži izmjenu starijih, nego da se naznače razlike i razriješe sukobi među njima.

6. Minimalističko oblikovanje - W3 Konzorcij ima namjeru standardizirati samo najpotrebnije tehnologije. Zamišljeno je da se time olakšava izrada aplikacija na već postojećim dobro definiranim standardima, dok se istovremeno istražuju buduće složenije tehnologije.

„Dobro napravljeni semantički web“ i stvarna, konkretna implementacija semantičkog weba predstavlja koncept povezanih podataka. Govoreći o konceptu povezanih podataka s tehničkog stajališta misli se na metodu stvaranja poveznica između pojedinih ‘dijelova’ RDF dokumenata koji opisuju izvore. Kako bi se to ostvarilo, svaki dio RDF dokumenta koji se želi povezati mora biti identificiran HTTP URI-jem. Iz toga možemo zaključiti da koncept povezanih

²² Usp. Farago, Filip.; Bosančić, Boris.; Badurina, Boris. Povezani podaci i knjižnice. Str. 30-31.

²³ *Uniform Resource Identifier* - kompaktan niz znakova za identificiranje apstraktnog ili fizičkog izvora/resursa.

podataka najviše počiva na dvjema glavnim tehnologijama semantičkog weba HTTP-u, URI-ju i RDF-u.²⁴

URI je najvažnija vrsta identifikatora za potrebe koncepta povezanih podataka. Njegova se sintaksa definira kao kompaktan niz znakova koji identificiraju apstraktni ili fizički izvor, tj. niz znakova koji se podvrgavaju sintaktičkim pravilima i koriste set znakova American Standard Code for Information Interchange (ASCII).²⁵ URI može koristiti razne sheme. Radi se o prefiksima koji stoje na početku URI-ja. Za njihovu registraciju zadužena je agencija IANA - The Internet Assigned Numbers Authority. Postoji pedesetak različitih shema, a neke od njih su http, ftp, mailto, telnet, gopher i druge. Za potrebe povezanih podataka, važna je jedino http shema te URI-ji temeljeni na njoj nazivaju se HTTP URI-ji. U sklopu specifikacije URI-ja, definira se i pojam izvora kao bilo što što se može identificirati URI-jem, te koji je važan za opis RDF-a. Bitno je naglasiti da URI jedinstveno identificira izvor, a ne reprezentaciju istog. Ako jedan izvor može imati više reprezentacija, na zahtjev klijenta, poslužitelj odabire odgovarajuću. To može biti dokument na jeziku korisnika, dokument prilagođen prikazu na mobilnim uređajima i sl.²⁶ Nadalje, RDF predstavlja podatkovni model koji se može izraziti kroz više strojno čitljivih jezika, a najčešći standard je RDF/XML. RDF zajedno s RDF Schemom (RDFS) predstavlja temeljni sloj semantičkog weba sa stajališta razmjene metapodataka na podatkovnoj razini (s pretpostavljenom serijalizacijom u XML-u). RDF izražava informacije pomoću jednostavnih izjava (engl. RDF statements). Njihova struktura sastoji se od tri dijela (RDF trojke): subjekta, predikata i objekta. Uzevši u obzir da URI-ji identificiraju bilo što kao izvor, subjekt i predikat RDF izjave moraju biti izvori identificirani URI-jima, dok objekt RDF izjave može, ali ne mora biti izvor identificiran URI-jem odnosno može biti i običan niz znakova. U praksi, to znači da ‘dijelovi’ RDF izjava koji su identificirani URI-jem mogu upućivati na ‘dijelove’ odnosno izvore drugih RDF izjava. Drugim riječima, objekt jedne RDF izjave može postati subjektom druge i tako se, barem u načelu, može ići unedogled.²⁷ Opisani postupak, u najkraćim crtama, sažima bit koncepta povezanih podataka. Konačno, skup međusobno povezanih RDF izjava čini jedan divovski globalni graf.²⁸

²⁴ Isto. Str. 31.

²⁵ Isto. Str. 32.

²⁶ Isto.

²⁷ Isto. Str. 32-33.

²⁸ Abele, Andrejs; McCrae, John P.; Buitelaar, Paul; Jentzsch, Anja; Cyganiak, Richard. Linking Open Data cloud diagram 2017. URL: <http://lod-cloud.net/> (2017-06-30)

3.1. Temeljna načela semantičkog weba i povezanih podataka

Tim Berners-Lee iznosi četiri komponente na kojima se gradi koncept povezanih podataka.

1. Za identificiranje pojedinog opisa mrežnog izvora ("stvari") koristi se URI/IRI (*Uniform/International Resource Identifier*)
2. Za prijenos URI/IRI-ja koristi se HTTP protokol
3. 'Stvari' su opisane u RDF/XML formatu
4. Opisi 'stvari' uključuju linkove (URIs/IRIs) prema drugim srodnim mrežnim izvorima.²⁹

Navedena načela mogu se svrstati u tri kategorije: jednostavnost, stabilnost i upravljivost. Jednostavnost je prisutna zbog kratkih 'mnemoničkih' URI-ja koji su lako pamtljivi i koje je lako proslijediti putem elektroničke pošte, a stabilnost je važna zbog što duže postojanosti URI-ja koji se postavio za neki izvor. Nadalje, preporuča se uključivanje godine u putanju, tako da ako dođe do moguće promjene na URI-ju iz te postavljene godine da to ne utječe na ostale godine.³⁰ Temeljni razlozi zbog kojih se koriste HTTP URI-ji su da oni predstavljaju jednostavan način za stvaranje jedinstvenih imena na globalnoj razini te da oni ne služe samo kako imena, nego i kao metoda pristupanja informacijama koje opisuju identificirani objekt.³¹

4. Bibliografski podaci kao povezani podaci

Važnost objavljivanja bibliografskih podataka knjižnice je u tome da su otvoreni i slobodno dostupni u obliku koji je zajednički, proširiv i kojeg se može lako ponovno koristiti. Knjižnice, uz svoje standarde, zagovaraju i pokušavaju usvojiti i današnje internetske standarde koji omogućuju ponovnu uporabljivost, umrežavanje i dijeljenje metapodataka.³² S obzirom na znatnu količinu ulaganja i napora koje su organizacije, kao što su IFLA, Kongresna knjižnica, OCLC i mnoge druge nacionalne knjižnice već proveli, na razvoju i održavanju MARC formata, prijelaz na povezane podatke nije trivijalan način i put usvajanja povezanih podataka ne može proći bez svojih izazova. Kada govorimo o knjižnicama pojavljuju se tri glavna izazova. Prvi izazov predstavlja opsežno korištenje MARC standarda koji služi kao osnova za postojeće sustave upravljanja knjižnicama. Danas doslovno postoji milijardu zapisa na MARC formatu te za svaki pokušaj da se napravi najmanji odmak od njega imao bi ogromne implikacije u smislu samih resursa koji su

²⁹ Usp. Farago, Filip.; Bosančić, Boris.; Badurina, Boris. Povezani podaci i knjižnice. Str. 33.

³⁰ Usp. Isto. Str. 34.

³¹ Usp. Isto.

³² Usp. Alemu, Getaneh; Stevens, Brett; Ross, Penny; Chandler, Jane. Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. // New Library World 113, 11/12(2012), str. 555. URL: <http://www.emeraldinsight.com/doi/pdfplus/10.1108/03074801211282920> (2017-07-08)

opisani.³³ Drugi je izazov terminološka razlika koja postoji između knjižničnih i internetskih standarda. U vezi s tim predlaže se jednostavni model koji se fokusira na objekt (predmet) kao što je specifična knjiga više nego na njegove apstrakcije kao što su djelo, izrazi i manifestacije. No takve apstrakcije ni nisu u svakodnevnoj upotrebi od strane knjižničara, izdavača, katalogizatora i korisnika knjižnice. Također, preporučuje se da knjižnice i zajednica povezanih podataka rade zajedno kako bi olakšale ponovnu uporabljivost i proširivost podataka knjižnice.³⁴ Treći i vrlo važan izazov s kojim se susreću potencijali usvojitelji je složenost tehnologije povezanih podataka kao što su RDF/XML, RDFS i OWL. Postoji očigledan nedostatak alata i aplikacija za stvaranje povezanih podataka u knjižnicama. Dakle, nužno je da takve tehnologije budu jednostavne za korištenje, no one su uglavnom previše komplicirane za osobe izvan zajednice povezanih podataka. Smatra se da osoba koja ima osnovne vještine u izradi mrežnih stranica je i sposobna stvoriti stranicu na temelju spomenutih tehnologija povezanih podataka.³⁵

4.1. Knjižnični zapis UNIMARC u okruženju semantičkog weba

Semantičko modeliranje može se svesti na tri koraka, a to su odabir i rastavljanje odabranog zapisa u pojedinačne RDF-izjave (postupak disagregacije), dodjeljivanje URI-ja sastavnim dijelovima disagregiranog zapisa (identifikatorima, poljima opisa, odnosno elementima metapodataka te vrijednostima elemenata metapodataka) te objavu zapisa u RDF/XML-u, pogodnom za dijeljenje u okruženju povezanih podataka.³⁶

Postupak kreće odabirom, u ovom slučaju, konkretnog bibliografskog zapisa u UNIMARC formatu. Preuzet je iz CROLIST-ova kataloga Filozofskog fakulteta u Osijeku (<http://crolist.ffos.hr/cgi-bin/unilib.cgi>) i prikazan u nastavku.

00000nam0#2200000# #45##

001			110719006
005			20150506143500.1
010	##	\$a	978-953-7036-91-1
100	##	\$a	20150323d2015#####y0hrvy0191####ba
101	0#	\$a	hrv
102	##	\$a	HR

³³ Usp. Isto. Str. 556.

³⁴ Usp. Isto. Str. 556-557.

³⁵ Usp. Isto. Str. 557.

³⁶ Usp. Farago, Filip; Bosančić, Boris; Badurina, Boris. Povezani podaci i UNIMARC: semantičko modeliranje knjižničnih zapisa UNIMARC u okruženju semantičkog weba. // Seminar Arhivi, knjižnice, muzeji 17, (2014), str. 152. URL: <http://hrcak.srce.hr/ojs/index.php/akm/article/viewFile/3566/2076> (2017-04-26)

		\$b	H
105	##	\$a	y#####000ad
200	1#	\$a	U zemlji Plavog Noja
		\$e	roman prostora uma
		\$f	Boris Bosančić
210	##	\$a	Zagreb
		\$c	Naklada Breza
		\$d	2015
215	##	\$a	275 str.
		\$d	21 cm
320	##	\$a	Inv. br.: 89.621sk
327	1#	\$a	Bilješka o piscu: str. 273
608	##	\$3	811228001
		\$a	Psihološki roman
		\$x	Hrvatska književnost
608	##	\$3	830508020
		\$a	Satirički roman
		\$x	Hrvatska književnost
675	##	\$a	821.163.42-31
700	#1	\$3	250502119
		\$a	Bosančić
		\$b	Boris
801	#0	\$a	HR
		\$b	GISKO
		\$g	HR PPIAK
801	#2	\$a	HR
		\$b	FFOSIJEK
		\$g	HR PPIAK
920	##	\$a	HR
		\$b	BP16-0324
981	##	\$z	bk1
990	##	\$a	886.2-3
		\$b	BOS
		\$c	u
991	##	\$a	p
		\$b	89.621sk
992	##	\$a	1sk
998	##	\$x	0515-IČ
		\$y	0515-IČ
		\$z	0515-IČ

Uočimo da se na početku zapisa nalazi broj 110719006, koji predstavlja identifikator koji ovaj kataloški zapis identificira u relacijskoj bazi podataka sustava CROLIST te taj broj postaje

subjektom svake RDF-izjave u RDF-dokumentu. Nadalje, disagregacija zapisa odnosi se na rastavljanje zapisa na pojedinačne RDF-izjave. U tablici (Tablica 1.) prikazat će se stvaranje RDF-tripleta (subjekt, predikat, objekt) od UNIMARC-potpolja konkretnog FFOS-ova kataloškog zapisa. Uz to, zamijenit će se postojeći identifikator novostvorenim HTTP URI-identifikatorom. Postojeći URL-identifikator (<http://crolist.ffos.hr/cgi-bin/unilib.cgi?form=D1110719006>) mogli bismo koristiti, no kako je već prethodno navedeno URI-ji moraju biti postojani i ne bi se trebali mijenjati. Dakle, navedeni URI ne posjeduje svojstva poput jednostavnosti, stabilnosti i upravljivosti te zaključujemo da ovaj URI nije pogodan za korištenje unutar mreže povezanih podataka. Za potrebe ovog primjera dodijelit će se jednostavniji HTTP URI: <http://katalog.ffos.hr/>, kojem ćemo pridružiti prefiks ffos za potrebe reprezentaciju u RDF/XML kodu. U ovom konkretnom primjeru URI bi glasio <http://katalog.ffos.hr/110719006>. U nastavku koristit ćemo prefiks kojeg smo prethodno dodijelili zbog jednostavnosti samog zapisa.³⁷

Tablica 1. Stvaranje RDF-tripleta od UNIMARC potpolja konkretnog FFOS-ova kataloškog zapisa s novostvorenim HTTP URI identifikatorom³⁸

URI (subjekt)	Svojstvo (predikat)	Vrijednost svojstva (objekt)
ffos:110719006	(ima) jedinstveni identifikator unutar sustava.	110719006
ffos:110719006	(ima) autora	Boris Bosančić
ffos:110719006	(ima) naslov	U zemlji Plavog Noja
ffos:110719006	(ima) mjesto izdanja	Zagreb
ffos:110719006	(ima) izdavača	Naklada Breza
ffos:110719006	(ima) ISBN	978-953-7036-91-1
ffos:110719006	(ima) godinu izdavanja	2015
ffos:110719006	(ima) broj stranica	275
ffos:110719006	(ima) dimenzije	21
ffos:110719006	(ima) jezik	hrvatski
ffos:110719006	(ima) predmetnice	Psihološki roman Hrvatska književnost
ffos:110719006	(ima) UDK oznaku	821.163.42-31
...

³⁷Usp. Isto. Str. 153-156.

³⁸ Usp. Isto. Str. 154-156.

Nadalje, u sljedećoj tablici (Tablica 2.) prikazat će se dodjeljivanje URI-ja predikatima i objektima RDF-izjave. RDF zahtijeva da se elementima i vrijednostima pridodaju strojno čitljivi identifikatori koji su u skladu sa sintaksom URI-ja.³⁹ Što se tiče dodjeljivanja URI-ja predikatima potražiti će se semantički istovrijedni elementi u drugim shemama metapodataka koje su objavljene kao povezani podaci (MARC21, RDA i sl.) ili u nekoj od ontologija (npr. BIBO).⁴⁰ Postupak dodjeljivanja URI-ja elementima metapodataka završava zamjenom naziva elemenata URI-ima, odnosno zamjenom polja zapisa s elementima metapodataka iz RDF-rječnika.⁴¹ U ovom primjeru koristit će se svojstva iz tri imenska prostora: Dublin Core (DCMI Metadata Terms, URI: <http://purl.org/dc/terms/>) (DC elements, URI: <http://purl.org/dc/elements/1.1/>), ISBD-a (ISBD elements, URI: <http://iflastandards.info/ns/isbd/elements/>) i BIBO-ontologije (The Bibliographic Ontology, URI: <http://purl.org/ontology/bibo/>).⁴² Na kraju, postupak modeliranja završavamo korakom koji se odnosi na pronalaženje URI-ja za objekte RDF-izjave. Vrlo važno je spomenuti da sami termini u načelu predstavljaju neki smisleni niz alfanumeričkih znakova, no kada je vrijednost objekta RDF-izjave navedena u obliku URI-ja, ona može postati subjektom drugog RDF-tripleta te u tom slučaju možemo reći da se koncept povezanih podataka ostvario.⁴³ Predmetne odrednice kao povezani podaci objavljeni su od Kongresne knjižnice (Library of Congress Subject Headings (LCSH)), Dewey Decimal Classification (DDC), UDC Summary Linked Data, FAST (Faceted Application of Subject Terminology), VIAF (Virtual International Authority File) i dr.⁴⁴ Od rječnika vrijednosti, u našem primjeru, odabrani su VIAF, LCSH i UDC što možemo vidjeti u sljedećoj tablici (Tablica 2.).

Tablica 2. Disagregirani kataloški zapis kao skup RDF-tripleta s naznačenim prefiksima imenskih prostora⁴⁵

URI (zapisa)	UNIMARC →	Element s prefiksom imenskog prostora i njegov URI	Vrijednost
ffos:110719006	001	dc:identifier http://purl.org/dc/elements/1.1/identifier	110719006

³⁹ Usp. Dunsire, Gordon; Willer, Mirna. UNIMARC and linked data. Str. 3. URL: <http://journals.sagepub.com/doi/abs/10.1177/0340035211430008> (2017-04-26)

⁴⁰ Usp. Usp. Farago, Filip; Bosančić, Boris; Badurina, Boris. Povezani podaci i UNIMARC. Str. 157.

⁴¹ Usp. Isto.

⁴² Usp. Isto.

⁴³ Usp. Isto. Str. 159-160.

⁴⁴ Usp. Isto.

⁴⁵ Usp. Isto. Str. 161.

ffos:110719006	700 \$a Osobno ime - Primarna intelektualna odgovornost	dc:creator http://purl.org/dc/terms/creator	http://viaf.org/viaf/305670443/#Bosančić, _Boris
ffos:110719006	200 \$a Glavni stvarni naslov	isbd:P1004 hasTitleProper http://iflastandards.info/ns/isbd/elements/P1004	U zemlji Plavog Noja
ffos:110719006	210 \$a Mjesto izdavanja	isbd:P1168 hasPlaceOfPublicationProductionDistributionStatement http://iflastandards.info/ns/isbd/elements/P1168	http://id.loc.gov/authorities/names/n79116205
ffos:110719006	210 \$c Nakladnik	dc:publisher http://purl.org/dc/elements/1.1/publisher	http://purl.org/dc/elements/1.1/publisher/naklada_breza
ffos:110719006	010 \$a Međunarodni standardni knjižni broj (ISBN)	bibo:isbn10 http://purl.org/ontology/bibo/isbn10	978-953-7036-91-1
ffos:110719006	200 \$d Godina izdavanja	isbd:P1018 hasDateOfPublicationProductionDistribution http://iflastandards.info/ns/isbd/elements/P1018	2015
ffos:110719006	215 \$a Posebna oznaka građe i opseg (broj stranica)	isbd:P1054 hasPageination http://iflastandards.info/ns/isbd/elements/P1054	275
ffos:110719006	215 \$d Dimenzije	isbd:P1024	21

		hasDimensions http://iflastandards.info/ns/isbd/elements/P1024	
ffos:110719006	101 \$a Jezik jedinice	dc:language language http://purl.org/dc/terms/ language	http://lexvo.org/id/is o639-3/hrv
ffos:110719006	606 \$x Predmetnica	dc:subject http://purl.org/dc/terms/subject	http://id.loc.gov/auth orities/subjects/sh85 034180
ffos:110719006	675 \$a Univerzalna decimalna klasifikacija (UDK)	dc:subject http://purl.org/dc/terms/subject	http://udcdata.info/0 67968

Slijedom razmatrane metodologije, pojedine RDF-izjave koje su iskazane tablično mogu se serijalizirati u RDF/XML-u. Dokument započinje deklaracijom upotrijebljenih imenskih prostora te se imenski prostor navodi samo u okviru taga <rdf:description> u koji se potom ugnježđuju svi ostali predikati i objekti RDF-izjava kako slijedi:

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:isbd="http://iflastandards.info/ns/isbd/elements/"
xmlns:bibo="http://purl.org/ontology/bibo/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
<rdf:Description rdf:about="http://katalog.ffos.hr/110719006">
  <dc:identifier>110719006</dc:identifier>
  <dc:creator rdf:resource="http://viaf.org/viaf/305670443" />
  <isbd:hasTitleProper>U zemlji Plavog Noja</isbd:hasTitleProper>
  <isbd:hasPlaceOfPublicationProductionDistributionStatement
rdf:resource="http://id.loc.gov/authorities/names/n79116205" />
  <dc:publisher rdf:resource="http://id.loc.gov/authorities/
names/naklada_breza"/>
  <bibo:isbn10>978-953-7036-91-1</bibo:isbn10>
```

```

<isbd:hasDateOfPublicationProductionDistribution>2015
</isbd:hasDateOfPublicationProductionDistribution>
<isbd:hasPagination>275</isbd:hasPagination>
<isbd:hasDimensions>21 cm</isbd:hasDimensions>
<dc:language rdf:resource="http://lexvo.org/id/iso639-3/hrv"/>
<dc:subject rdf:resource="http://id.loc.gov/authorities/subjects/sh85034180" />
<dc:subject rdf:resource="http://udcdata.info/067968" />
</rdf:Description>
</rdf:RDF>

```

Konverzija UNIMARC zapisa u RDF nije nimalo jednostavan proces zbog bogate strukture formata UNIMARC koji zahtijeva pronalaženje elemenata iz nekoliko različitih imenskih prostora. Time se dovodi u opasnost semantička konzistentnost novog zapisa u odnosu na izvorni zapis. Optimalno rješenje bilo bi korištenje imenskog prostora UNIMARC-a.⁴⁶

4.2. Problemi primjene koncepta povezanih podataka na knjižničnim zapisima

Glavni problemi objave knjižničnih bibliografskih podataka kao povezanih podataka se dijele na nekoliko dijelova:

1. nepostojanje RDF rječnika
2. nekompatibilnost s mrežnom tehnologijom i manjkavost podataka
3. organizacijski problemi
4. tehnološki problemi.⁴⁷

Metapodaci korišteni od strane knjižnice nemaju svoju RDF reprezentaciju te važne metapodatkovne komponente nisu identificirane URI-jima. Navedeni organizacijski problemi uključuju nedostatak otvorenih licenci koje omogućuju ponovnu uporabu povezanih podataka te nepostojanje suglasnosti oko idealnog paketa ponovno upotrebljivih RDF paketa za one aplikacije koje koriste povezane podatke knjižnica. Uz nedostatak tih otvorenih licenci, postoji i nedostatak onih koje bi obuhvatile API-je⁴⁸, klijentski softver i podatkovne standarde u okruženju povezanih podataka.⁴⁹ Glavni problem za pretvaranje postojećih zapisa u povezane podatke predstavlja

⁴⁶ Usp. Isto. Str. 164.

⁴⁷ Usp. Farago, Filip.; Bosančić, Boris.; Badurina, Boris. Povezani podaci i knjižnice. Str. 40.

⁴⁸ API - Application programming interface (sučelje za programiranje aplikacija)

⁴⁹ Usp. Farago, Filip.; Bosančić, Boris.; Badurina, Boris. Povezani podaci i knjižnice. Str. 40.

odabir rječnika. Kako bi se skup podataka objavio kao povezani podaci mogu se konstruirati vlastiti rječnici ili se mogu koristiti postojeći, no preporuča se korištenje općeprihvaćenih rječnika. Nadalje, sa stajališta UNIMARC-a, odnosno kataloga koji ga koriste, najveći problem je što ne postoji reprezentacija istoga u RDF-u, odnosno, ne postoji imenski prostor u kojem bi URI-ji za svaki element UNIMARC-a bili definirani te uz to ne postoji ni rječnik koji opisuje njihove odnose.⁵⁰ Zbog toga, postojana su dva smjera za RDF reprezentaciju kataloga utemeljenih na UNIMARC-u, a to su izrada cjelovite ontologije UNIMARC-a koristeći RDF Scheme ili OWL⁵¹ i mapiranje pojedinih elemenata koristeći više različitih semantički najbližih rječnika. Najsrodniji rječnik za UNIMARC je ISBD čija se RDF reprezentacija nalazi u registru RDF rječnika i elemenata metapodataka – OMR (Open Metadata Registry).⁵² Trenutno je moguće izvršiti prebacivanje UNIMARC zapisa u neki drugi format na sintaktičkoj razini bez gubljenja izvornih podataka, a to nam omogućuje mehanizam XML Scheme koji omogućuje formaliziranje strukture XML dokumenta. Zaključno, tehnologija povezanih podataka, sa stajališta primjene u knjižnicama, stvara probleme jer ne postoje dobro utvrđena pravila, tj. dobri primjeri iz prakse.⁵³ Neke od preporuka Stalnom odboru UNIMARC-a za daljnju raspravu i odobrenje su odobriti metodu identificiranja UNIMARC elemenata i rječnika, odlučiti o početnom stvaranju i održavanju UNIMARC elemenata i rječnika u OMR-u, podržavati i promicati prijevod elemenata i svojstava UNIMARC-a na nacionalne jezike, razmotriti zahtjeve koji se odnose na aplikacijske profile za UNIMARC, istražiti daljnje širenje UNIMARC svojstava i pratiti razvoj MARC21-a, posebno izmjenu njegovog bibliografskog okvira.⁵⁴

⁵⁰ Usp. Isto. Str. 41.

⁵¹ Web Ontology Language

⁵² Usp. Farago, Filip.; Bosančić, Boris.; Badurina, Boris. Povezani podaci i knjižnice. Str. 41.

⁵³ Usp. Isto. Str. 41-42.

⁵⁴ Usp. Dunsire, Gordon; Willer, Mirna. Nav. dj. Str. 16.

Zaključak

Katalogiziranje, metapodaci i povezani podaci su međusobno različiti, ali povezani. Katalogiziranje se brzo mijenja i postaje sve složenije kako se odmiče od MARC-a i AACR2 do RDA, RDF-a, povezanih podataka i semantičkog weba. Cilj ovog završnog rada bio je prikazati kako je međunarodni format nastao te kako su knjižnični zapisi, strojno čitljivi u formatu UNIMARC, prikazani kao povezani podaci. Svrha povezanih podataka je, kako i sam naziv kaže, povezati različite izvore te svakog pojedinog jedinstveno identificirati uz pomoć URI-ja. To se nastojalo prikazati kroz tri poglavlja u kojima se definirao strojno čitljiv format UNIMARC, analizirala glavna i temeljna načela semantičkog weba i povezanih podataka te prikazalo semantičko modeliranje knjižničnog zapisa u okruženju povezanih podataka. Može se zaključiti da nije nimalo jednostavno milijarde zapisa na UNIMARC formatu usvojiti kao povezane podatke te da se svi knjižničari i drugi sudionici u semantičkom modeliranju knjižničnih zapisa susreću i susretat će se u budućnosti s velikim izazovima. Od izazova koji uključuju kompliciranu tehnologiju koja iz dana u dan sve više napreduje, katalogizatori se moraju suočiti i s izazovom terminološke prirode između knjižničnih i internetskih standarda. Veliku složenost konverzije UNIMARC zapisa u RDF čini pronalaženje elemenata iz različitih imenskih prostora što nas dovodi do toga da je najbolje rješenje za to kreirati i koristiti jedinstveni imenski prostor UNIMARC formata. Katalogizatori moraju poznavati svaki format i sva pravila koja se primjenjuju jer moraju održavati naslijeđene podatke, ali i krenuti naprijed s novim vrstama podataka.

Literatura

Dunsire, Gordon; Willer, Mirna. UNIMARC and linked data. // IFLA Journal 37, 4(2011), str. 1-18. URL: <http://journals.sagepub.com/doi/abs/10.1177/0340035211430008> (2017-04-26)

Farago, Filip; Bosančić, Boris; Badurina, Boris. Povezani podaci i UNIMARC: semantičko modeliranje knjižničnih zapisa UNIMARC u okruženju semantičkog weba. // Seminar Arhivi, knjižnice, muzeji 17, (2014), str. 145-168. URL: <http://hrcak.srce.hr/ojs/index.php/akm/article/viewFile/3566/2076> (2017-04-26)

Farago, Filip; Bosančić, Boris; Badurina, Boris. Povezani podaci i knjižnice. // Vjesnik bibliotekara Hrvatske 56, 4(2013), str. 25-52. URL: <http://hrcak.srce.hr/142376> (2017-03-25)

Frederick, E. Donna. Metadata specialists in transition: from MARC cataloging to linked data and BIBFRAME (data deluge column). // Library Hi Tech News 33, 4(2016), str. 1-5. URL: <http://www.emeraldinsight.com/doi/abs/10.1108/LHTN-03-2016-0015> (2017-07-08)

Alemu, Getaneh; Stevens, Brett; Ross, Penny; Chandler, Jane. Linked Data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data models. // New Library World 113, 11/12(2012), str. 549-570. URL: <http://www.emeraldinsight.com/doi/pdfplus/10.1108/03074801211282920> (2017-07-08)

Abele, Andrejs; McCrae, John P.; Buitelaar, Paul; Jentzsch, Anja; Cyganiak, Richard. Linking Open Data cloud diagram 2017. URL: <http://lod-cloud.net/> (2017-06-30)

Rubin, E. Richard. Foundations of Library and Information Science. New York; London: Neal-Schuman Publishers, 2004.

Sarić, Ivana. Model povezivanja i interoperabilnosti knjižničnih zapisa na semantičkom webu. // Vjesnik bibliotekara Hrvatske 56, 1/2(2013), str. 107-122. URL: <http://hrcak.srce.hr/file/170267> (2017-04-26)

UNIMARC. // Hrvatska enciklopedija. Leksikografski zavod Miroslav Krleža. URL: <http://www.enciklopedija.hr/natuknica.aspx?ID=63215> (2017-04-26)

UNIMARC: bibliografski format. Zagreb: Hrvatsko knjižničarsko društvo; Zadar: Sveučilište, 2009.

Willer, Mirna. UNIMARC u teoriji i praksi. Rijeka: Naklada „Benja“, 1996.